BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 7: e111338
doi: 10.3897/biss.7.111338

OPEN ACCESS

Conference Abstract

A Strategy for Building the WFO Plant List

Roger Hyam‡, Alan Elliott‡

‡ Royal Botanic Garden Edinburgh, Edinburgh, United Kingdom

Corresponding author: Roger Hyam (rhyam@rbge.org.uk)

Received: 18 Aug 2023 | Published: 18 Aug 2023

Citation: Hyam R, Elliott A (2023) A Strategy for Building the WFO Plant List. Biodiversity Information Science and Standards 7: e111338. https://doi.org/10.3897/biss.7.111338

Abstract

All scientists will face the challenge of explaining what they do to a friend or relative. Fortunately it is easy for us to explain our work. We are building a list of all known plants. Unfortunately this elicits the awkward question: Hasn't that been done already? Everyone knows that Linnaeus started the naming convention in the 18th century. Surely we would have created a list of species in the intervening 270 years. Alas, there is no single, global species list. In 2022, when the team at the Royal Botanic Garden Edinburgh (RBGE) took on the coordination of the World Flora Online (WFO) Plant List, we considered what we could do differently to save our successors from this awkward dinner party question.

The WFO Plant List's primary purpose is as a structure for the WFO information portal. The portal contains a large amount of information. The list is a simple database of names and their taxonomic statuses. It currently contains 1.52 million names and 440,000 accepted taxa. Because the list has a global scope and includes all vascular plants and bryophytes, it has great potential to be of use outside the WFO portal. Functions might include a:

- common vocabulary for ecological monitoring networks;
- drop down list in a garden management system;
- destination for taxonomic output beyond a monographic paper;
- bridge from historical, observational studies to contemporary, molecular, phylogenetic research.

In short, the WFO Plant List can be a single, shared lookup table for plant taxa.

© Hyam R, Elliott A. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

There are four well known elements of project management: resources, timescale, quality and scope. We have limited control over the first three of these elements. For resources, our institutes have committed a part of our salaried time to facilitate the project but the vast amount of the work has to be done through collaboration with others. We can only inspire people to contribute and this must be done through principles of FAIR (Findable, Accessible, Interoperable and Reusable) data discussed below. There is no natural timescale for our work; we have therefore established a somewhat artificial drum beat of twice-yearly data releases. This enables us to prioritise smaller batches of work. In a list like this, quality is synonymous with accuracy and non-negotiable. If we have an error in our list, it must be corrected. The only element we have full control over is scope. We can choose what is included and what is not. We do this through the design of our data model. The simpler we can make the model, the more complete we can make the list and the easier it will be to improve quality.

We only include names that appear effectively/validly published under the International Code of Nomenclature for Algae, Fungi and Plants (ICNAFP). This is an explicit set of rules we can use to enforce data integrity. Unlike the Catalogue of Life, the Global Biodiversity Information Facility (GBIF) or the Global Names Architecture (GNA), we do not have to model names governed by other nomenclatural codes and can focus our resources. From the start, we have separated nomenclature from taxonomy. This gives us a clear set of nomenclatural facts supported by appropriate references that will not change over time, alongside taxonomic opinion that is linked to relevant supporting literature. We only support a single consensus taxonomy but by keeping snapshots of the taxonomy every six months, we allow changes in the science to be tracked through time. The separation of nomenclature from taxonomy within our identifier schema allows third parties to maintain their own classifications whilst mapping to our classification through taxonomically neutral name identifiers.

If we had been working a decade or more ago, we would have created tables for ancillary data such as literature, specimens and people. Today we can take advantage of the many data sources available via web links and only store data on nomenclatural acts and taxonomic placement. All other data is represented by a generic referencing mechanism. A reference consists of a URL (including digital object identifiers (DOIs) in URL form) and a citation string. This approach dramatically increases our ability to focus on taxonomic coverage and leaves specialist systems such as International Plant Names Index (IPNI), Biodiversity Heritage Library (BHL) and WikiData to handle other classes of data.

More important than the way we model the data is how it is produced and consumed by others. As a node in a graph of linked biodiversity information, our success is measured by the number of links we have to other nodes and people.

The data is being produced and maintained by a growing community organised into Taxonomic Expert Networks (TENs). There are about 300 individual scientists in 44 approved TENs. These TENs can contribute to the live dataset via submission of bulk data or by using a dedicated editing platform called Rhakhis. Care is taken to give attribution for contributions at the finest level of granularity possible using Open Researcher Contributor

Identifiers (ORCID). We strive to have the data available in bulk and at the level of each name under FAIR principles. All data is released under a Creative Commons CC0 licence. It is made available through the WFO portal, a dedicated API, ChecklistBank and Zenodo on a six-monthly release cycle. The dataset has a citable DOI as well as each version having its own DOI. All names have a stable URI and each version of each taxon has a stable URI. There is a name-to-ID matching service available through the API and as a web interface, and there are two R packages (WorldFlora and wfor) to facilitate analysis workflows.

Keywords

taxonomy, nomenclature, API, name matching, botanical code, ICBN, ICNAFP, checklist, World Flora Online

Presenting author

Roger Hyam

Presented at

TDWG 2023

Hosting institution

Royal Botanic Garden Edinburgh

Conflicts of interest

The authors have declared that no competing interests exist.